

Review of single clustering methods

Nurshazwani Muhamad Mahfuz¹, Marina Yusoff², Zakiah Ahmad³

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

²Advanced Analytic Engineering Center (AAEC), Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

³Institute for Infrastructure Engineering and Sustainability Management, Faculty of Civil Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Article Info

Article history:

Received May 2, 2019

Revised Jun 12, 2019

Accepted Jul 25, 2019

Keywords:

Clustering

Hierarchical method

Partition method

Single clustering

Unsupervised learning

ABSTRACT

Clustering provides a prime important role as an unsupervised learning method in data analytics to assist many real-world problems such as image segmentation, object recognition or information retrieval. It is often an issue of difficulty for traditional clustering technique due to non-optimal result exist because of the presence of outliers and noise data. This review paper provides a review of single clustering methods that were applied in various domains. The aim is to see the potential suitable applications and aspect of improvement of the methods. Three categories of single clustering methods were suggested, and it would be beneficial to the researcher to see the clustering aspects as well as to determine the requirement for clustering method for an employment based on the state of the art of the previous research findings.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Marina Yusoff,

Advanced Analytic Engineering Center (AAEC),

Faculty of Computer and Mathematical Sciences,

Universiti Teknologi MARA,

Shah Alam, Selangor, Malaysia.

Email: marinay@tmsk.uitm.edu.my

1. INTRODUCTION

Identification of homogeneous group is an exploratory cluster analysis has been researched since few decades ago [1-5]. Various applications used clustering methods as such of neural network [6-7], artificial intelligence [8-11] and statistics [12-16]. There are several clustering methods that have been enhanced to cater the issues of dataset features such as big data sets [17], high-dimensional [18] and distributed data sets [19]. Clustering can be divided into two such as traditional technique includes hierarchy based, partition based, fuzzy based, distribution based, density based, graph based, grid based, fractal based and model based and modern techniques include kernel based, ensemble based, swarm intelligence based, quantum theory based, spectral graph theory based, affinity propagation based, distance and density based, spatial, data based, data stream based and large scale data based [20-21]. There are several issues of traditional clustering technique that arises in the study such as when the partition set of patterns to be clustered is large or the features reaches higher dimensions. Thus, the judgement would be difficult to justify, obtain and take time to execute [22-24].

Traditionally clustering techniques are broadly used are hierarchical, partitioning, grid based and density based clustering analysis [25-26]. Partitioning clustering such as k-means are useful to determine the number of clusters in the data. However, each pattern of mostly traditional clustering techniques which belongs to one cluster and provides a single view [27-28]. The importance view and less importance view of data clustering are treated equally, therefore, there is a lack to distinguish to compare. Traditional clustering can perform poorly when the non-optimal result exist in terms of between-cluster heterogeneity and highly

feature dimension due to the presence of multicollinearity among the variables or high skewness of the clustering variables [29]. This paper addresses the review of traditional single clustering methods. The remainder of this article is organized as follows: In Section 2, we present the approaches in methods for clustering in terms of single clustering solution. Section 3 is a discussion and a highlights drawback of single clustering methods and finally presents a conclusion.

2. SINGLE CLUSTERING TECHNIQUES

Many previous studies that used different clustering methods in a single and multiple solution for a given problem of classification, image segmentation and pattern recognition. Merging similar items and separating dissimilar items in different groups and only provides one single solution known as single clustering. This review divides the single clustering into three categories, namely hierarchical methods, partition method and other clustering methods. Hierarchical methods and partition methods are the most used in many domains as demonstrated in Figure 1. The following subsections discuss each category.

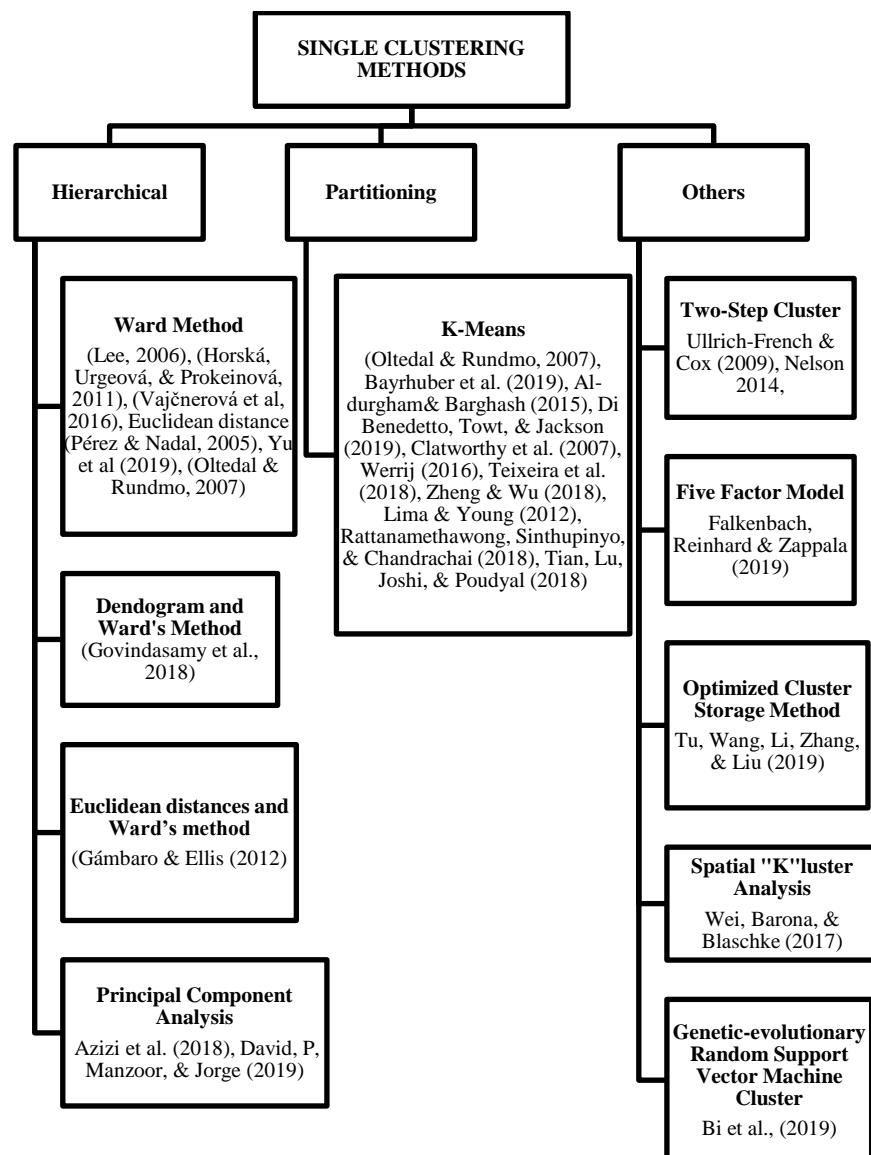


Figure 1. Single clustering methods

2.1. Hierarchical clustering methods

Hierarchical clustering as a recursive partitioning perform clustering from various types of data sets. It can be classified into agglomerative and divisive. Agglomerative clustering method is normally known as bottom-up approach where each data point represents a single cluster at the beginning and these clusters will be merged based on the similarities and all points belong to the same cluster considering the most two similar cluster [30]. Meanwhile, divisive is a top-down approach is the inverse of agglomerative clustering. Initially, all points belong to the same cluster, which assigned all of the observations to single cluster and then partition the cluster into two least similar clusters and proceed on each cluster until there is one cluster for each observation [31-32]. Various research on hierarchical methods for clustering analysis were done include the application of Ward hierarchical to find subgroups regarding self-reported physical and mental health status of patients [33] and applied hierarchical cluster analysis to investigate the impact of parenting styles with adolescent self-esteem of middle school student in Korean culture. In addition, the used of cluster analysis to examine the effect of residents' attitudes in the Balearic Islands of Spain regard tourism on community for five different opinion groups [34] meanwhile the investigated service quality in a metropolitan public bus service quality of Granada using cluster analysis and a Pearson correlation in order to discover the variables that give effect to the passengers' about the service [35]. Those hierarchical methods provide solutions for clustering different types of datasets, however, it is still a lack of accuracy and require more datasets for a good clustering solution.

For instance, hierarchical cluster analysis and expectation maximum (EM) algorithm were applied based on Gaussian mixture model (EMGM) in order to evaluate the mass spectrometry of the soluble species of Shengli lignite into meaningful groups [36] still need improvement in terms of results [12]. Another application that used hierarchical clustering is to obtain the personality of similarity on different group of individuals and the characteristics of cultural perceived the transport risks. In this case, the Ward method as a cluster analysis and principal component analysis were implemented in evaluating the quality of a tourist destination [15]. Additionally, cluster analysis was used in predicting the cluster of Mid-Atlantic wine market based on purchasing behavior, attitudes and social demographic attributes [16]. Various domains of problem performed by Principal Component Analysis (PCA) such as to upgrade the consumption data of electricity and the result showed that it can decrease the time of examining as well as increases the accuracy of clustering and applied in identifying control, heavy metal variability in soil and dust at children's playgrounds and reflect potential sources in a different group of location [37].

A method of hierarchical cluster analysis with Euclidean distances and Ward's aggregation method were performed in classifying the groups of consumers with different perceptions towards the different types of chocolate healthiness [38]. Hierarchical clustering is applied to identify the segmentation of tourist visitors of the chosen destinations on memorable tourism experience rates [39]. Correlation analysis, principal component analysis (PCA) and cluster analysis (CA) were used to investigate air pollution for the concentrations of nitrogen oxides, ozone, and particulate matter [40]. Ward methods as clustering technique was applied in defining consumer perception on food choice and quality in different selected Central European countries such as Poland, the Slovak republic and the Czech Republic [41]. Despite all findings from the mentioned above, all methods are feasible for the clustering results, however a lot of improvement when consider large data sets, numbers of variables and its variety. It is a challenge when to solve big data solutions.

2.2. Partition clustering methods

Partitioning clustering method occurs when each of the object must belong to exactly one group and each group must contain at least one object. Partitioning clustering method occurs when each of the object must belong to exactly one group and each group must contain at least one object [42]. A very popular K-means methods falls under this category. K-means clustering were used in many domain applications includes identifying patients' perception of the different quality perspectives [43], to examine the clusters from self-care behaviors associated with sleep quality and mental health and to the cluster perception of illness patients as well as in a hybridation with other method [57]. The work of [2] compared the method of clustering such as Average Linkage, Complete Linkage, K-means clustering to examine the perception of illness and the findings revealed that K-means was given the better and appropriate result for illness research. In other research, the method of Rank-Order clustering algorithm was compared with the K-Means and Spectral in order to investigate the run-time complexity and cluster quality in terms of external and internal of face label and the findings revealed the performances of Rank-Order clustering algorithm is better than K-Means and Spectral [44]. K-Means and Spectral clustering were applied to calculate the distance between two clusters of ordinal survey data in a highly structured ranking, and the results indicated that the combination of K-Means and Spearman rank-order coefficient outperforms in the least of executions of time [45]. Physiological signals of human personalized stress used k-means and regression neural network

were evaluated where the finding indicates that the accuracy was improved compared to traditional technique without clustering [46]. K-Means cluster analysis was applied to interpret the sources, temporal and spatial trends of ultrafine particles (UFP). Latent Dirichlet Allocation and the VSM model with K-Means clustering were applied to calculate the similarity of text for a cluster analysis, then the conclusion shown that the accuracy of clustering algorithms is enhanced [47]. Euclidean Distance K-means method was used to evaluate the public perception of social and environmental impacts of hydropower projects [14]. Fuzzy clustering and K-means clustering techniques were compared to the creation of perfectionism profiles, then for the fuzzy clustering, the similarity depends on the quantity of cluster overlap [48].

Traditional K-Means with improved algorithm such as K-Prototypes and improved K-Prototypes algorithm in evaluating Intrusion Detection System and the results revealed that improved algorithm can achieve a higher detection rate and lower false alarm rate than the traditional K-Means [49]. Logistic regression and the K-Mean clustering technique were used in clusters alumni into segments in terms of the characteristics, lifestyles, types of behavior, and interests of alumni [50]. K-means clustering techniques was evaluated by examining the attitudes, opinions and interests in forest certification by segmentation of respondents of a landowner survey in Shandong, China [51].

2.3. Other single clustering methods

This section will briefly elaborate the application of other single clustering methods in which are not categorized under both hierarchical and partition methods. As demonstrated in Figure 1, five methods include a two step cluster analysis, five factor model, optimized cluster storage method, Spatial “K”luster analysis (SKATER), and Genetic-evolutionary Random Support Vector Machine Cluster will be focused. A two-step cluster analysis, which is profiled analysis was implemented in order to study the variation the characteristics in different student segments with gender as a categorical variable and task value of students’ perceptions, self-efficacy and attitudes on the usage and applications of computers [52]. The two step cluster analysis was applied to examine the differences of profile analysis in relative levels of self-determination and consequences of physical education motivation [53]. An optimized cluster storage method was proposed in order to cut the utilization and cost of data storage in the file for real-time big data in the Internet of Things, and the results showed the effectiveness and 70% of storage cost less than traditional methods meanwhile SKATER, fuzzy clustering and Gaussian mixture modelling for model-based clustering, and the result shows SKATER technique of individual perception effectively reflect the internal heterogeneities of citizens’ perceptions across multiple scales [54].

Genetic-evolutionary Random Support Vector Machine Cluster (GERSVMC) algorithm was proposed to evaluate the accuracy for Autism Spectrum Disorder (ASD) patients and the results indicates that GERSVMC is sufficient with 96.8% of accuracy [9]. Clustering with high dimensional Gaussian mixtures was proposed based on the expectation-maximization (EM) algorithm and a direct estimation method for the sparse discriminant vector, then the findings revealed that the proposed method shows better performance compared to conventional low-dimensional in terms of the optimal rate of convergence. Five Factor Model as a cluster model was applied based to find subtypes of personality traits [55].

3. DISCUSSION

The traditional clustering techniques are broadly studied in few decades ago. It only considers one single way to partition data into groups. The aim of clustering is to identify group or clusters that are like each other than to other clusters. There are several advantages of single clustering which are simple as well as ease of handling and understanding of any types of similarity and straightforward. Eventhough traditional clustering technique have been useful, it is quite challenging to handle situations as sich when clustering processes in the population of overlapping or ambiguous. However, k-means work well when the shape of the clusters is hyper spherical. K-means also requires prior knowledge of clusters. The drawback of traditional cluster analysis as stated in the following:

- Limited of space and time complexity
Traditional clustering method are expensive in terms computational and storage requirements. The storage required for the hierarchical clustering technique is very high when the number of data points are high as need to store the similarity matrix in the random-access memory. The time is limited as there is a need to perform numbers iterations and, in each iteration, need to update the similarity matrix. Therefore, it seems not suitable for for large data.
- Population overlap
Generally, cluster made up of neighbouring elements, therefore, the elements within a cluster tend to be homogeneous. The situation where the elements appears simultaneously in more than one cluster

called overlapping cluster analysis. The redundancy of information occurs if the cluster analysis overlapped. Therefore, there was an inaccurate description of data if use traditional clustering.

- Limited of information knowledge
Most of traditional clustering cannot capture variety in various application due to interpretation of traditional clustering analysis was too strict since it can only interpret set of groups. For instance, grouping topic or region in a one cluster in a news and webpages [56].
- Low performance accuracy
Most of essential parameters of conventional clustering are determined by the human user's experience and cannot obtain balance between accuracy and efficiency of the clustering process [58]. The accuracy performance of traditional clustering becomes low since importance view and less important view of respondents are treated equally. For example, misinterpretation of the hierarchical analysis such as dendrogram.

4. CONCLUSIONS

Clustering is an unsupervised machine learning and can be used to improve the accuracy of supervised machine learning algorithms by clustering the data points into similar groups. From the reviews, many single clustering methods were applied in many real-world problems such as transport, finance, behavioral and health. It seems to offer a feasible solution, however, the accuracy is still low and require more computational resources. Some drawback of the methods is briefly discussed. One of the important aspects of clustering such as outliers and insufficient population needs to cater even though clustering is easy to implement in a various number of domains has to be considered for an improvement of the methods or a new method to be constructed.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Ministry of Higher Education (Fundamental Research Grant Scheme (FRGS) Grant: 600-IRMI/FRGS 5/3 (213/2019)) and Universiti Teknologi MARA for the financial support of this research project.

REFERENCES

- [1] P. Murena and U. M. R. Mia-paris, "An Information Theory based Approach to Multisource Clustering," pp. 2581–2587, 2007.
- [2] J. Clatworthy, M. Hankins, D. Buick, J. Weinman, and R. Horne, "Cluster analysis in illness perception research: A Monte Carlo study to identify the most appropriate method," *Psychol. Heal.*, vol. 22, no. 2, pp. 123–142, 2007.
- [3] T. S. Madhulatha, "An Overview Clustering Method," *IOSR J. Eng.*, 2013.
- [4] M. Garza-Fabre, J. Handl, and J. Knowles, "An Improved and More Scalable Evolutionary Approach to Multiobjective Clustering," *IEEE Trans. Evol. Comput.*, vol. 22, no. 4, pp. 515–535, 2018.
- [5] J. Nasiri and F. M. Khiyabani, "A whale optimization algorithm (WOA) approach for clustering," *Cogent Math. Stat.*, vol. 5, no. 1, pp. 1–13, 2018.
- [6] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *ImageNet Classification with Deep Convolutional Neural Networks*, 2012.
- [8] S. Dutta, A. K. Das, G. Dutta, and M. Gupta, *Emerging Technologies in Data Mining*. Springer Singapore, 2010.
- [9] X. Bi *et al.*, "The Genetic-evolutionary Random Support Vector Machine Cluster Analysis in Autism Spectrum Disorder," *IEEE Access*, vol. PP, no. c, pp. 1–1, 2019.
- [10] V. Tunali, T. Bilgin, and A. Camurcu, "An Improved Clustering Algorithm for Text Mining: Multi-cluster Spherical K-means," *Int. Arab J. Inf. Technol.*, vol. 13, no. 1, pp. 12–19, 2016.
- [11] G. Chao, S. Sun, and J. Bi, "A Survey on Multi-View Clustering," pp. 1–17, 2018.
- [12] S. Olstedal and T. Rundmo, "Using cluster analysis to test the cultural theory of risk perception," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 10, no. 3, pp. 254–262, 2007.
- [13] M. Bumgardner, G. W. Graham, P. C. Goebel, and R. Romig, "Perceptions of Firms within a Cluster Regarding the Cluster's Function and Success: Amish Furniture Manufacturing in Ohio," *Proc. 17th Cent. Hardwood For. Conf.*, vol. 78, no. 740, pp. 597–606, 2011.
- [14] G. R. Lima and C. E. F. Young, "Cluster analysis as a tool to assess the public perception of social and environmental impacts of hydropower projects," 2012.
- [15] I. Vajčnerová, J. Šácha, K. Ryglová, and P. Žižan, "Using the cluster analysis and the principal component analysis in evaluating the quality of a destination," *Acta Univ. Agric. Silv. Mendelianae Brun.*, vol. 64, no. 2, pp. 677–682, 2016.

- [16] R. Govindasamy, S. Arumugam, J. Zhuang, K. M. Kelley, and I. Vellangany, "Cluster Analysis of Wine Market Segmentation – A Consumer Based Study in the Mid-Atlantic USA," vol. 63, no. 1, pp. 151–157, 2018.
- [17] E. Azizi, H. K. Shishavan, B. Mohammadi-Ivatloo, and A. M. Shotorbani, "Clustering Electricity Big Data for Consumption Modeling Using Comparative Strainer Method for High Accuracy Attainment and Dimensionality Reduction," *J. Energy Manag. Technol.*, vol. 3, 2018.
- [18] T. T. Cai, J. Ma, and L. Zhang, "CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality," *Ann. Stat.*, vol. 47, no. 3, pp. 1234–1267, Jun. 2019.
- [19] Q. Tong, X. Li, and B. Yuan, "Efficient distributed clustering using boundary information," *Neurocomputing*, vol. 275, pp. 2355–2366, Jan. 2018.
- [20] S. Singh Ghuman, "Clustering Techniques-A Review," *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, no. 5, pp. 524–530, 2016.
- [21] R. Joshi and P. Mulay, "Mind Map based Survey of Conventional and Recent Clustering Algorithms: Learning's for Development of Parallel and Distributed Clustering Algorithms," *Int. J. Comput. Appl.*, vol. 181, no. 4, pp. 14–21, 2018.
- [22] M. Steinbach, L. Ertoz, and V. Kumar, "of Clustering High Dimensional Data," 2004.
- [23] C. Bouveyron, S. Girard, and C. Schmid, "High Dimensional Data Clustering."
- [24] M. Pavithra and R. M. S. Parvathi, "A Survey on Clustering High Dimensional Data Techniques," vol. 12, no. 11, pp. 2893–2899, 2017.
- [25] P. Berkhin, "Survey of clustering data mining techniques," *Tech. Report, Accrue Softw.*, pp. 1–56, 2002.
- [26] C. Budayan, I. Dikmen, and M. T. Birgonul, "Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11772–11781, 2009.
- [27] A. Jain, M. Murty, and P. Flynn, "Data Clustering : A Review," vol. 31, no. 3, 1999.
- [28] B. G. O. Reddy, "Literature Survey On Clustering Techniques," *IOSR J. Comput. Eng.*, vol. 3, no. 1, pp. 01–12, 2012.
- [29] J. F. Hair and W. C. Black, "Cluster Analysis," pp. 147–206, 2000.
- [30] C. Diffusion, "CHAPTER Diffusion and Contagion," no. c, 2005.
- [31] A. Rana and A. Sharma, "Analysis of Clustering Techniques in Data Mining," no. May, pp. 36–39, 2018.
- [32] A. V. Jiman and P. H. K. Khanuja, "Literature Survey: Clustering Technique," *Int. J. Comput. Appl. Technol. Res.*, vol. 5, no. 10, pp. 671–674, 2016.
- [33] M. Bayrhuber *et al.*, "Perceived health of patients with common variable immunodeficiency – a cluster analysis," *Clin. Exp. Immunol.*, pp. 1–10, 2019.
- [34] E. A. Pérez and J. R. Nadal, "Host community perceptions. A cluster analysis," *Ann. Tour. Res.*, vol. 32, no. 4, pp. 925–941, 2005.
- [35] R. de Oña, G. López, F. J. D. de los Rios, and J. de Oña, "Cluster Analysis for Diminishing Heterogeneous Opinions of Service Quality Public Transport Passengers," *Procedia - Soc. Behav. Sci.*, vol. 162, no. Panam, pp. 459–466, 2014.
- [36] Y. R. Yu *et al.*, "Mass spectrometric evaluation of the soluble species of Shengli lignite using cluster analysis methods," *Fuel*, vol. 236, no. May 2018, pp. 1037–1042, 2019.
- [37] Y. Jin, D. O'Connor, Y. S. Ok, D. C. W. Tsang, A. Liu, and D. Hou, "Assessment of sources of heavy metals in soil and dust at children's playgrounds in Beijing using GIS and multivariate statistical analysis," *Environ. Int.*, vol. 124, no. January, pp. 320–328, 2019.
- [38] A. Gámbaro and A. C. Ellis, "Exploring consumer perception about the different types of chocolate," *Brazilian J. food technology*, vol. 15, no. 4, pp. 317–324, 2012.
- [39] S. B. Miftarević, "Tourist Segmentation in Memorable," pp. 110–119, 2018.
- [40] N. David, L. V. P. S. Manzoor, and O. C. Jorge, "Statistical Tools for Air Pollution Assessment : Multivariate and spatial analysis studies in the madrid region," vol. 2019, 2019.
- [41] E. Horská, J. Urgeová, and R. Prokešínová, "Consumers' food choice and quality perception: Comparative analysis of selected Central European countries," *Agric. Econ.*, vol. 2011, pp. 493–499, 2011.
- [42] S. S. J and S. Pandya, "An Overview of Partitioning Algorithms in Clustering Techniques," vol. 5, no. 6, pp. 1943–1946, 2016.
- [43] M. Di Benedetto, C. J. Towt, and M. L. Jackson, "A Cluster Analysis of Sleep Quality, Self-Care Behaviors, and Mental Health Risk in Australian University Students," *Behav. Sleep Med.*, vol. 00, no. 00, pp. 1–12, 2019.
- [44] C. Otto, D. Wang, and A. K. Jain, "Clustering Millions of Faces by Identity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 289–303, 2018.
- [45] P. Werrij, "Clustering Ordinal Survey Data in a Highly Structured Ranking," 2016.
- [46] Q. Xu, T. Nwe, and C. Guan, "Cluster-based analysis for personalized stress evaluation using physiological signals," *IEEE J. Biomed. Heal. ...*, vol. 19, no. 1, pp. 275–281, 2015.
- [47] N. Zheng and J. Y. Wu, "Cluster Analysis for Internet Public Sentiment in Universities by Combining Methods," *Int. J. Recent Contrib. from Eng. Sci. IT*, vol. 6, no. 3, p. 60, 2018.
- [48] J. H. Bolin, J. M. Edwards, W. Holmes Finch, and J. C. Cassidy, "Applications of cluster analysis to the creation of perfectionism profiles: A comparison of two clustering approaches," *Front. Psychol.*, vol. 5, no. APR, pp. 1–9, 2014.
- [49] L. P. Su and J. A. Zhang, "The Improvement of Cluster Analysis in Intrusion Detection System," *Proc. - 10th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2017*, vol. 2017-Octob, pp. 274–277, 2017.

- [50] N. Rattanamethawong, S. Sinthupinyo, and A. Chandrachai, "An innovation model of alumni relationship management: Alumni segmentation analysis," *Kasetsart J. Soc. Sci.*, vol. 39, no. 1, pp. 150–160, 2018.
- [51] N. Tian, F. Lu, O. Joshi, and N. C. Poudyal, "Segmenting landowners of Shandong, China based on their attitudes towards forest certification," *Forests*, vol. 9, no. 6, pp. 1–14, 2018.
- [52] K. Nelson, "Student motivational profiles in an introductory MIS course : An exploratory cluster analysis," *Res. High. Educ. J.*, vol. 24, pp. 1–14, 2014.
- [53] S. Ullrich-French and A. Cox, "Using Cluster Analysis to Examine the Combinations of Motivation Regulations of Physical Education Students," *J. Sport Exerc. Psychol.*, vol. 31, no. 3, pp. 358–379, 2009.
- [54] L. Tu, Y. Wang, P. Li, C. Zhang, and S. Liu, "An optimized cluster storage method for real-time big data in Internet of Things," *J. Supercomput.*, no. 0123456789, 2019.
- [55] D. M. Falkenbach, E. E. Reinhard, and M. Zappala, "Identifying Psychopathy Subtypes Using a Broader Model of Personality: An Investigation of the Five Factor Model Using Model-Based Cluster Analysis," *J. Interpers. Violence*, p. 088626051983138, 2019.
- [56] D. Niu, J. G. Dy, and Z. Ghahramani, "A Nonparametric Bayesian Model for Multiple Clustering with Overlapping Feature Views," *Proc. 15th Int. Conf. Artif. Intell. Stat.*, vol. XX, pp. 814–822, 2012.
- [57] Verma, I., Kaur, A., & Kaur, I. Refined Clustering of Software Components by Using K-Mean and Neural Network. *IAES International Journal of Artificial Intelligence*, 4(2), 62, 2015.
- [58] L. Liao, K. Li, K. Li, C. Yang, and Q. Tian, "A multiple kernel density clustering algorithm for incomplete datasets in bioinformatics," *BMC Syst. Biol.*, vol. 12, no. Suppl 6, 2018.

BIOGRAPHIES OF AUTHORS



Nurshawani Muhamad Mahfuz is currently a PhD student in Information Technology at Universiti Teknologi MARA (UiTM) Shah Alam. Her education background consist of a Bachelor of Science (Statistics) (Hons) at Universiti Teknologi MARA (UiTM), Malaysia. She continued her education on Master of Applied Statistics at Universiti Teknologi MARA (UiTM), Malaysia. Her current research interests include data mining, clustering, multi-view learning, artificial intelligent, optimization and their application in timber.



Marina Yusoff is currently a Deputy Dean (Research and Industry Linkages) and Senior Lecturer of Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, Malaysia. She holds a PhD in Intelligent System from the Universiti Teknologi MARA. She previously worked as a senior executive of Information Technology in SIRIM Berhad, Malaysia. She holds a bachelor's degree in computer science from the University of Science Malaysia, and Master of Science in Information Technology from Universiti Teknologi MARA. She is interested in the development of intelligent application, modification and enhancement computational intelligence techniques include particle swarm optimisation, neural network, genetic algorithm, and ant colony. She has many impact journals publications and presented her research in many conferences locally and internationally.



Zakiah Ahmad is currently a Dean Faculty of Civil Engineering Universiti Teknologi MARA (UiTM), Malaysia. She received a PhD in Timber Engineering from University of Bath, United Kingdom. She received Master in Mathematics and bachelor in civil engineering at University of Memphis, Memphis, United States. She is currently a professor and director of Institute for Infrastructure Engineering and Sustainable Management (IIESM) at UiTM Shah Alam, Malaysia. Her current research interest includes timber engineering, timber adhesive composite, wood or fibre cement composites and glued-in rod.